

DOCUMENT RESUME

ED 322 167

TM 015 284

AUTHOR Lopez, Kay A.
TITLE Testing Interaction Effects without Discarding Variance.
PUB DATE Nov 89
NOTE 25p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (Little Rock, AR, November 8-10, 1989).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Analysis of Variance; *Aptitude Treatment Interaction; Comparative Analysis; Methods Research; *Multiple Regression Analysis

ABSTRACT

Analysis of variance (ANOVA) and multiple regression are two of the most commonly used methods of data analysis in behavioral science research. Although ANOVA was intended for use with experimental designs, educational researchers have used ANOVA extensively in aptitude-treatment interaction (ATI) research. This practice tends to make researchers prone to categorizing continuous variables. Categorizing variables in a non-experimental design using an ANOVA analysis produces in both investigators and their audience/clients the illusion that the investigator has experimental control over the independent variable. Multiple regression prevents the researcher from being lulled into this trap by allowing variables to retain their highest level of scale and requiring that the researcher examine data from several different perspectives. A hypothetical data set is presented to illustrate the superiority of multiple regression in analyzing data from ATI experiments. Multiple regression not only does all that ANOVA can, it also provides more explanatory power. Six data tables are included. (TJH)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it.
☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official
OERI position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

KAY A. LOPEZ

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

ED322167

TESTING INTERACTION EFFECTS WITHOUT DISCARDING VARIANCE

Kay A. Lopez

University of New Orleans

and

Louisiana State University Medical Center

BEST COPY AVAILABLE

Paper presented at the annual meeting of the Mid-South
Educational Research Association, Little Rock, November 10,
1989.

ABSTRACT

ANOVA and multiple regression are two of the most commonly utilized methods of data analysis in behavioral science research. Although ANOVA was intended for use with experimental designs, educational researchers have used ANOVA extensively in aptitude-treatment interaction (ATI) research. A hypothetical data set is presented to illustrate the superiority of multiple regression in analyzing data from ATI experiments. Multiple regression not only does all that ANOVA can, it also provides more explanatory power.

Variance is the heart of scientific explanation (Kerlinger, 1986). Without the ability to account for sufficient variance, a psychometric instrument cannot be considered reliable (Thompson, 1986). Science requires that the researcher look at data as sets of "ordered pairs" and to determine to what extent these pairs covary (Kerlinger, 1986, p.56).

It is widely recognized today that phenomena studied in behavioral science are complex in the sense that they are best explained by more than one predictor in relation to criterion measures (Kerlinger & Pedhazur, 1973). ANOVA and multiple regression are two commonly utilized methods of data analysis in educational research (Elmore & Woehlke, 1988; Gaither & Glorfeld, 1985; Goodwin & Goodwin, 1985). This paper discusses the merits of both these statistical models in relation to the aptitude-treatment interaction (ATI) designs so common to educational research.

Method 1: ANOVA

ANOVA, developed by Fisher in 1925, was devised for use in experimental research designs. ANOVA has been widely used in educational and biomedical research where the effects of one or more treatments are evaluated in relation to a criterion measure. For example, a researcher might want to examine the impact of amount of exercise (high, moderate,

low), diet (moderate versus low fat), and relaxation training (experimental versus control group) in relation to variations in diastolic blood pressure. In this example, each of the predictors is manipulated by the researcher and is truly experimental in nature.

In order to use ANOVA, predictors must be nominal in scale and uncorrelated with each other. This requirement is generally met by assigning equal numbers of subjects to cells of the design and by ensuring that all possible combinations of experimental conditions are represented.

Unfortunately, ANOVA has been widely used in education and psychology to analyze data from studies with both experimental and nonexperimental predictors. This type of design is known as the classic aptitude-treatment interaction (ATI) model advocated by Cronbach (1957). For example, the researcher who is interested in intelligence (high, average, low), field independence (high, medium, low), and teaching strategy (traditional versus nontraditional) in relation to test scores is utilizing an ATI design. Teaching strategy is the experimental variable; the others are known as "attribute" or "measured" variables because they describe characteristics inherent to subjects (Kerlinger & Pedhazur, 1973, p.7).

Attribute variables are often intercorrelated; for

example, high IQ subjects may tend to be more field independent than low IQ subjects. Collinearity among variables makes it difficult for the researcher to meet ANOVA requirements that an equal number of subjects be present in all cells of the design or even that all cells be represented (Kerlinger & Pedhazur, 1973). The design described above represents a 3x3x2 ANOVA, which requires that 18 cells be filled with equal numbers of subjects. It can be anticipated that the researcher would have difficulty finding subjects to represent all the discrete combinations needed. Moreover, for practical or ethical reasons, this type of assignment may not be feasible in educational research.

Also, most psychometric instruments that measure attribute variables utilize interval measurement scales. Researchers who wish to use ANOVA must convert interally scaled attribute variables to nominal categories (e.g., high, medium, low). This forces the predictors to be uncorrelated. It also distorts the natural relationships among predictors and between predictors and the dependent variable. Finally, the conversion discards variance, thus adversely affecting reliability of measurement (Cohen, 1968). This sacrificing of variance was called a "cardinal sin in research" by Thompson (1988a, p. 3) and can lead to erroneous research conclusions. It also distorts the complexity of educational reality.

Table 1 presents a hypothetical data set for a 2x2 ANOVA. The predictors are field independence (FI) and teaching strategy (TCH); the dependent variable is test scores (DV). Table 1 also presents the intervally scaled FI variable dichotomized (FIGRP) into "high" versus "low" categories to meet ANOVA requirements. The computer commands for creating FIGRP are presented in the Appendix. The variable TCH involves two levels: traditional teaching strategy (1) and nontraditional teaching strategy (2).

Scrutiny of the data set reveals that while high FI subjects obtain their highest scores in the nontraditional teaching condition, FI is positively correlated overall with test performance. Table 2 presents the correlation matrix for the original variables and the dichotomized attribute variable FIGRP. Note that the bivariate r between FI in its original interval form and DV is .50, whereas the correlation between FIGRP and DV is only .01! This example illustrates the extent to which converting an intervally scaled variable to the nominal level of scale distorts reality by reducing natural variance. Low FI subjects have scores ranging from 2-9; high FI subjects have scores ranging from 10-18. By categorizing these subjects into two discrete groups, the differences between individual scores is ignored. As Cliff (1987, p. 30) notes,

Such division are not infallible. Think of the persons near the borders. Some who should be highs are actually classified as lows, and vice versa. In addition, the "barely highs" are classified the same as the "very highs," even though they are different. Therefore, reducing a reliable variable to a dichotomy makes the variable more unreliable, not less.

Thus, Pedhazur (1982, p. 454-453) argues that

Categorization of attribute variables is all too frequently resorted to in the social sciences...It is possible that some of the conflicting evidence in the research literature of a given area may be attributed to the practice of categorization of continuous variables...Categorization leads to a loss of information, and consequently, to a less sensitive analysis.

One reason why ANOVA has been popular with researchers is that discrete effects sizes for each predictor can be computed. The effect size is determined by dividing the sum of squares (SOS) for each predictor by the SOS total and converting this proportion to a percentage (SOS effect/SOS total X 100). Table 3 presents the ANOVA analysis of the data set. The effect sizes for FIGRP and TCH are .01 and

.13%, respectively, and these main effects do not achieve statistical significance. The interaction effect, FIGRPXTCH, is significant ($p < .001$) and the effect size is quite large indeed (95.27%).

Method 2: Multiple Regression

While the ANOVA results are easier to conceptualize initially than is the case with multiple regression, this paper illustrates that a multiple regression analysis of the same data set gives more accurate information and increases the probability of achieving statistical significance. In respect to multiple regression analysis, Kerlinger and Pedhazur (1973, p. 3) noted long ago that

It can be used equally well in experimental or non-experimental research. It can handle continuous and categorical variables. It can handle two, three, four or more independent variables... Finally, as we will abundantly show, multiple regression analysis can do anything the analysis of variance does--sums of squares, mean squares, F ratios-and more.

Multiple regression has several advantages when compared to ANOVA. Multiple regression does not require that predictors be converted to the nominal level of scale. Moreover, the researcher need not be concerned with the

issues of ANOVA cells and numbers of subjects in cells. Multiple regression can be used with both experimental and nonexperimental variables; these features make it a reliable and flexible method of data analysis (Darlington, 1968).

Interpreting the results of a multiple regression analysis requires a different approach than does ANOVA. Since predictors are not forced to be uncorrelated, discrete effect sizes for each predictor cannot be computed. Instead, the regression table partitions the sum of squares into SOS regression (explained) and SOS residual (unexplained). The multiple R statistic represents the extent to which all of the predictors contribute to variance in the dependent variable. Multiple R square is the effect size and represents the proportion of variance accounted for by the combination of the predictors (Shavelson, 1988). The null hypothesis tested relates to whether R square is equal to zero.

With multiple regression analysis, the individual effects of predictors can be evaluated in several ways. When predictors are uncorrelated, beta weights can give information as to the relative contribution of a predictor to variance in the dependent variable scores. When predictors are correlated, beta weights are unreliable, since they are computed based on the unique contribution of each

predictor and overlook shared variance between predictors (Kerlinger, 1986). Another way to assess the effect of a specific predictor is to compute a structure coefficient. A structure coefficient is the bivariate correlation between a predictor and the total explained variance in the dependent variable independent of the influence of other predictors in the regression equation (Thompson & Borello, 1985). It is computed by dividing the correlation between a specific predictor and the dependent variable by multiple R ($r_{x,y/R}$).

In summary, the researcher evaluating the results of a multiple regression analysis looks at 4 pieces of information:

- (a) statistical significance (is R square equal to zero?);
- (b) beta weights;
- (c) bivariate correlations between predictors; and
- (d) structure coefficients.

Some researchers erroneously believe that interaction effects and polynomial data trends can only be analyzed by ANOVA. Thompson (1989) demonstrated that both linear and curvilinear main and interaction effects can be studied using multiple regression.

The analysis of a linear interaction can be done by computing what is known as a "product variable" (Kerlinger & Pedhazur, 1973, p. 414). To create the product variable,

the predictors are converted to Z scores and multiplied times each other. The computer commands to create the Z scores (ZFI, ZTCH) are located in the Appendix. The product variable is given a unique name (in this case, INTER) and is entered into the regression equation as a distinct predictor along with FI and TCH.

Table 4 presents the multiple regression analysis of the data without the product variable. The effect size is moderate ($R^2 = 25.29\%$) and $p < .1503$, which indicates that the results are not statistically significant. Note, however, that FI has a sizable beta weight of .50 and a structure coefficient of 1.00, indicating a perfect correlation with explained variance in the dependent variable. Also, since statistical significance is mainly a function of sample size (Thompson, 1988b), the thoughtful researcher will note that if the sample size had been increased to $n=24$, F_{calc} would be increased to 3.55, which indicates statistical significance for the variable FI ($p < .05$). However, these results are not complete, since the interaction is not analyzed.

Table 5 presents the multiple regression analysis with the use of the product variable. In this case R^2 has increased to 79.05% and the results are statistically significant ($p < .0002$). The discrepancy between this effect

size and that of ANOVA (95.42%) is accounted for by the increased accuracy of multiple regression in computing effect sizes because variables retain their original level of scale (Thompson, 1986).

Table 8 presents the correlation matrix with information needed to compute the structure coefficients for the predictors. The structure coefficients for INTER, FI and TCH are .95, .56, and .04, respectively. The structure coefficients and the beta weights, in this example, give similar information as to the relative contribution of each predictor to variance in the dependent variable.

Both ANOVA and multiple regression inform us, in this hypothetical data set, that the main source of explained variance is accounted for by the interaction of field independence and teaching strategy. Note, however, that in the multiple regression analysis we also discover that the main effect of field independence also contributes significantly to the variance in test scores. This additional information emerges because the variable FI is allowed to retain its original interval scale. This example illustrates that ANOVA decreases power against Type II error and therefore decreases chances for attaining statistical significance.

Discussion

It is important for educational researchers to make

informed choices in respect to methods of analysis of their data. This especially true when practice or decision making will be affected by research outcomes (Thompson, 1988). In order to do justice to the complexity of educational phenomena, sophisticated and accurate methods of data analysis are needed. One principle that should be emphasized is that variance in data should never be discarded or minimized.

One reason why researchers may be prone to categorizing continuous variables is that some researchers unconsciously associate ANOVA with the power of experimental designs. As Thompson, (1981, p. 5) notes,

Researchers tend to prefer experimental over correlational research designs because experimental designs provide more complete information about causality. Why does this situation contribute to OVAism? Because some researchers confuse research designs with the statistical techniques which are used to analyze the data which the designs help to generate.

Humphreys (1978, p. 873) notes that:

The basic fact is that a measure of individual difference is not an independent variable, and it does not become one by categorizing the scores and treating the categories as if they defined a

variable under experimental control in a
factorially designed analysis of variance.

Similarly, Humphreys and Fleishman (1974, p. 468) note that categorizing variables in a nonexperimental design using an ANOVA analysis "not infrequently produces in both the investigator and his audience the illusion that he has experimental control over the independent variable. Nothing could be more wrong."

Multiple regression honors these precepts by allowing variables to retain their highest level of scale and requires that the researcher thoughtfully examine data from several different perspectives. When looking at options for analyzing data, researchers would do well to consider multiple regression as an alternative to ANOVA.

References

- Cliff, N. (1987). Analyzing multivariate data. San Diego: Harcourt Brace Jovanovich.
- Cohen, J. (1968). Multiple regression as a general data analytic system. Psychological Bulletin, 70, 426-443.
- Cronbach, L. J. (1957). The two disciplines of scientific psychology. American Psychologist, 12, 671-684.
- Darlington, R. B. (1968). Multiple regression in psychological research and practice. Psychological Bulletin, 69, 161-182.
- Elmore, P. G., & Woehkle, P. L. (1988, April). Research methods employed in American Educational Research Journal, Educational Researcher, and Review of Educational Research from 1978 to 1987. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.
- Fisher, R. A. (1925). Statistical methods for research workers. Edinburgh, England: Oliver and Boyd.
- Gaither, N., & Glorfeld, L. (1985). An evaluation of the uses of tests of significance in organizational behavior research. Academy of Management Review, 10, 787-793.
- Goodwin, L.D., & Goodwin, W. L. (1985). Statistical techniques in AERJ articles, 1979-1983: The preparation of graduate students to read the educational research literature. Educational Researcher, 14(2), 5-11.

- Humphreys, L. G. (1978). Doing research the hard way: Substituting analysis of variance for a problem in correlational analysis. Journal of Educational Psychology, 70, 873-876.
- Humphreys, L. G., & Fleishman, A. (1974). Pseudo-orthogonal and other analysis of variance designs involving individual difference variables. Journal of Educational Psychology, 66, 464-472.
- Kerlinger, F. N. (1986). Foundations of behavioral research (3rd ed.). New York: Holt, Rinehart and Winston.
- Kerlinger, F. N., & Pedhazur, E. J. (1973). Multiple regression in behavioral research. New York: Holt, Rinehart and Winston.
- Pedhazur, E. J. (1982). Multiple regression in behavioral research. New York: Holt, Rinehart and Winston.
- Shavelson, R. J. (1988). Statistical reasoning for the behavioral sciences. (2nd ed.). Boston: Allyn and Bacon.
- Thompson, B. (1981, November). The problem of OVAism. Paper presented at the annual meeting of the Mid-South Educational Research Association, Lexington KY. [Order document #03980, National Auxillary Publication Service, P.O. Box 3513, Grand Central Station, New York, NY 70017].
- Thompson, B. (1986). ANOVA versus regression analysis of ATI designs: An empirical investigation. Educational and Psychological Measurement, 46, 917-928.

- Thompson, B. (1988a). Discarding variance: A cardinal sin in research. Measurement and Evaluation in Counseling and Development, 21, 3-4.
- Thompson, B. (1988b, November). Common methodology mistakes in dissertations: Improving dissertation quality. Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville, KY. [ERIC Document Reproduction Service No. Ed 301 595].
- Thompson, B. (1989, January). Heuristics for the understanding of interaction, polynomial data trend, and the general linear model. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston, TX [ERIC Document Reproduction Service No. ED 304 453].
- Thompson, B. & Borello, G. M. (1985). The importance of structure coefficients in regression research. Educational and Psychological Measurement, 45, 203-209.

Table 1
Hypothetical Data Set for 2X2 ANOVA

ANOVA Cells		Variables			

		FI	FIGRP	TCH	DV
1:1		6	1	1	90
		7	1	1	92
		8	1	1	93
		9	1	1	94

1:2		2	1	2	70
		3	1	2	72
		4	1	2	74
		5	1	2	75

2:1		10	2	1	69
		11	2	1	71
		12	2	1	73
		13	2	1	76

2:2		15	2	2	89
		16	2	2	93
		17	2	2	95
		18	2	2	96

Table 2
Correlation Matrix: Original Variables
and Dichotomized Variable FIGRP

	FI	FIGRP	TCH	DV
FI	1.00			
FIGRP	.86	1.00		
TCH	.05	.00	1.00	
DV	.50	.01	.04	1.00

Table 3

Classical ANOVA

Source	Sum of Squares	df	Mean Square	F Calc	Sig of F	Effect Size
FIGRP	.25	1	.25	.04	.85	0.01%
TCH	2.25	1	2.25	.34	.57	.13%
FIXTCH	1640.25	1	1640.25	249.15	.00	95.27%
Residual	79.00	12	6.58			
Total	1721.75	15	114.78			

Table 4
Multiple Regression without Product Variable

Multiple R .50286
R Square .25287 (X 100)=25.29%

Source	Sum of Squares	df	Mean Square	F Calc	Sig of F
Regression	435.38	2	217.69	2.30	.1503
Residual	1286.37	13	98.95		
Total	1721.75	15	114.78		

Variables in the Equation	beta weight	Structure Coefficient
TCH	.01	(.04/.50)=0.08
FI	.50	(.50/.50)=1.00

Table 5

Multiple Regression with Product Variable

Multiple R .8891

R Square .79054 (X 100)=79.05%

Source	Sum of Squares	df	Mean Square	F Calc	Sig of F
Regression	1361.11	3	453.70	15.10	.0002
Residual	360.64	12	30.05		
Total	1721.75	15	114.78		

Variables in the Equation	beta weight	Structure Coefficient
INTER	1.83	(.85/.89)=.95
TCH	.06	(.04/.89)=.04
FI	-.43	(.50/.89)=.56

Table 6

Correlation Matrix: Original Variables and Product Variable

	FI	TCH	INTER	DV
FI	1.00			
TCH	.05	1.00		
INTER	.78	.00	1.00	
DV	.50	.04	.85	1.00

Appendix: SPSSX Computer Commands

```
1 Title "ANOVA vs Regression"
2 File Handle Int1/Name="Int1.dat"
3 Data List File=Int1/FI 1-2 TCH 3 DV 4-5
4 Subtitle "#1 Dichotomize Predictor FI for Classical ANOVA"
5 If (FI LT 10) FIGRP=1
6 If (FI GE 10) FIGRP=2
7 List Variables=all/cases=16
8 Subtitle "#2 Classical ANOVA"
9 ANOVA DV by FIGRP (1,2) TCH ( 1,2)
10 Correlations FIGRP TCH DV FI
11 Subtitle "#3 Multiple Regression with Original Variables"
12 Regression Variables=FI TCH DV/Descriptives=all/
13 Dependent=DV/Enter FI TCH
14 Subtitle "#4 Create Z Scores for Product Variable"
15 Compute ZFI=(FI-9.70)/5.11
16 Compute ZTCH=(TCH-1.5)/.52
17 Subtitle "#5 Create Product Variable"
18 Compute INTER=ZFI*ZTCH
19 Subtitle "#6 Multiple Regression with Product Variable"
20 Regression Variables=FI TCH DV INTER/Descriptives=all/
21 Dependent=DV/Enter FI TCH INTER
```